

STRIELAD - A Scalable Toolkit for Real-time Interactive Exploration of Large Atmospheric Datasets

Simon Schneegans¹ *

Lori Neary² †

Markus Flatken¹ ‡

Andreas Gerndt¹ §

¹German Aerospace Center (DLR), Dept. Simulation and Software Technology, ²Institut royal d'Aronomie Spatiale de Belgique

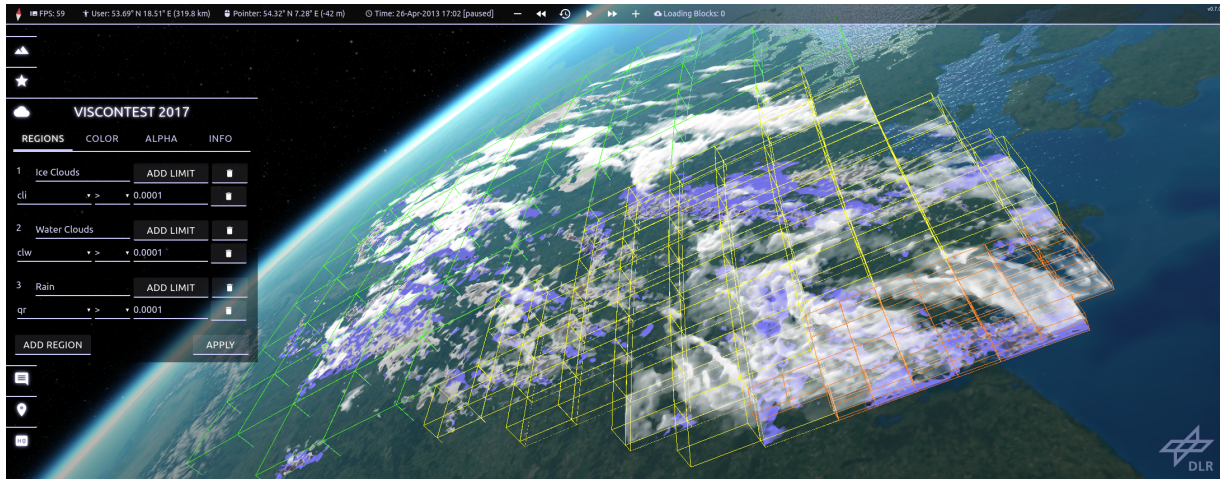


Figure 1: Real-time volumetric visualization of ice, water, and rain bands using our sub-volume selection mechanism in the graphical user interface. In addition, large terrain data is combined with the weather simulation to analyze their correlations.

ABSTRACT

Technological advances in high performance computing and maturing physical models allow scientists to simulate weather and climate evolutions with an increasing accuracy. While this improved accuracy allows us to explore complex dynamical interactions within such physical systems, inconceivable a few years ago, it also results in grand challenges regarding the data visualization and analytics process.

We present STRIELAD, a scalable weather analytics toolkit, which allows for interactive exploration and real-time visualization of such large scale datasets. It combines parallel and distributed feature extraction using high-performance computing resources with smart level-of-detail rendering methods to assure interactivity during the complete analysis process.

Index Terms: C.2.4 [Distributed Systems]: Client/Server—Distributed applications; I.3.2 [Graphics Systems]: Distributed/network graphics—Remote systems; I.6.6 [Simulation Output Analysis]: — [J.2]: Physical sciences and engineering—Earth and atmospheric sciences

1 MOTIVATION

The substantial progress in high-performance computing (HPC) together with an improved understanding of natural processes enable scientists to model and simulate physical phenomena with an

increasing accuracy and complexity. A major drawback of these technological advances is the growing gap between the possibilities to simulate and to analyze the produced data. It is not only a challenge of data complexity; the sheer amount of data, generated e.g. by numerical simulations, require tremendous effort to enable efficient data analysis and knowledge discovery. Especially interactive and explorative approaches are promising but also challenging. Yet they are important for urgent decision making in various fields such as traffic- or catastrophe management. Only when these gaps are closed, the full potential for knowledge discovery can be exploited.

STRIELAD, a scalable toolkit for real-time interactive exploration of large atmospheric datasets, will enable end-users to explore such large volumes of simulation data in real-time. It combines smart data acceleration structures, parallel and distributed processing, innovative data selection and interaction mechanisms. This is demonstrated with the 2.9 TB weather simulation dataset provided for the 2017 IEEE visualization contest. Additionally, this data is combined with various other data sources such as flight trajectories and geographical information obtained from satellite imagery using DLR’s terrain rendering software [7]. The goal of this paper is not to answer all given questions of the visualization contest. It is rather a description of STRIELAD, which in turn can be used to find those answers by real-time exploration.

2 SCALABLE VISUALIZATION TOOLKIT

A lot of applications for scientific visualization of simulation data are available on market. Some of them are scalable to enable the visualization of large-scale data [2] [4]. Furthermore, these applications often provide an extensive feature set for various application domains. However, to our knowledge none of them enables spatio-temporal exploration of huge simulation data with interactive frame rates and feature extraction in range of milliseconds.

With STRIELAD, end-users are able to interactively select sub-volumes of a dataset and apply boolean operations on them. For

*e-mail: simon.schneegans@dlr.de

†e-mail: lori.neary@aeronomie.be

‡e-mail: markus.flatken@dlr.de

§e-mail: andreas.gerndt@dlr.de

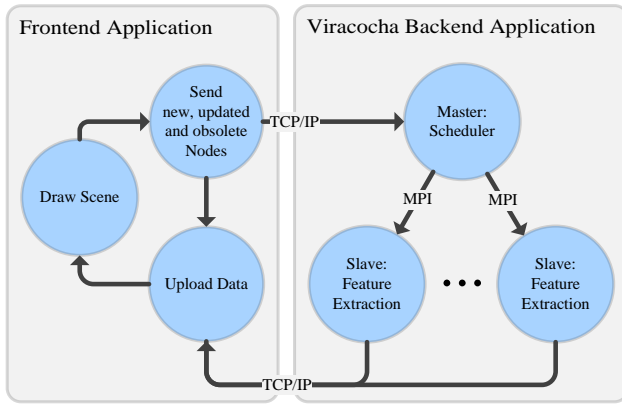


Figure 2: Distributed visualization setup: The frontend determines nodes for parallel processing. The backend executes the feature extraction on a HPC cluster system and streams results back.

each sub-volume, several scalar thresholds or domain limits can be defined. To speed-up feature extraction, the processing is importance driven, distributed and highly parallel. Combined with level-of-detail rendering and immediate streaming of result data, this creates a very responsive and interactive system. A shader editor in the graphical user interface (GUI) allows to control the appearance of the defined sub-volumes. The shading possibilities range from informative to realistic.

Figure 2 depicts the high-level architecture of the framework. A frontend application performs real-time rendering and user interaction. The GUI enables end-users to define multiple volumes of interests (depicted in Fig. 1). These sub-volumes are specified by several lower and upper scalar limits. Based on this information, the user’s position and the simulation time, requests are compiled and sent to a backend application. There, feature extraction is performed on an octree which has been generated in a pre-processing step. Furthermore, the user has the ability to modify shader code at runtime via the GUI. This allows for powerful coloring and shading of the received geometry.

2.1 Preprocessing

Before data can be explored with STRIELAD, a preprocessing step generates the required acceleration structure. In this step, the rectilinear input dataset is partitioned into many equally sized blocks (leaf nodes), and an octree is generated by sub-sampling (see Fig. 3). The additional memory requirement for the inner nodes of the octree is approximately 15 % of the size of the original data. This is about 400 GB for the visualization contest dataset.

Since several terabytes of data are read and written during the preprocessing, the actual execution time strongly depends on disc speed and may take several hours.

While the size of the individual octree nodes has no influence on the memory requirements of the acceleration structure, it significantly influences the visualization performance. If the size is small, many nodes are generated. As a consequence, feature extraction in the backend is highly parallel and visualization updates are received with low latency. However, since many nodes have to be rendered by the frontend, a lot of draw calls have to be issued and therefore the communication with the graphics driver may become a bottleneck.

The ideal node size depends on disc speed, backend processing power, network speed, CPU and GPU performance of the frontend. We achieved very good results with a node size of $20 \times 20 \times 20$ cells. This results in seven octree levels and a total node count of 51.393 per time step.

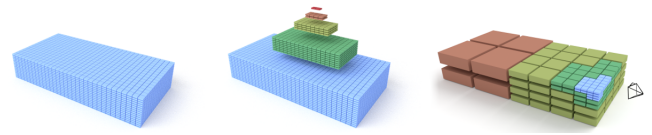


Figure 3: The original dataset is partitioned into many small data blocks (left). Then, by combining eight blocks into one, an octree is generated (center). When rendering, a set of nodes (called the ‘cut’) is chosen so that blocks close to the virtual camera have a high resolution (right).

2.2 Parallel feature extraction

The parallel and distributed feature extraction application is based on Viracocha [5]. As depicted in Fig. 2, our application implements a master-slave parallelization approach. The master process receives user queries from a frontend application, dissipates the requests into small independent work items, and distributes these work items to n slave instances for parallel processing. Furthermore, an integrated caching approach increases the I/O bandwidth for repeated data processing. We enhanced the scheduling approach of Viracocha to update the processing order of a running computation. This allows to fully change the processing order, e.g. discard previously visible data, insert new items for processing, or rearrange the data processing based on priority values.

The slave processes execute a boolean intersection algorithm for sub-volume extraction. The algorithm uses VTK [6] to load and process the individual octree nodes. It gets a description of pipelined intersection \cap operations defined by the limits per sub-volume. First, the dataset is clipped according to each limit, then the result is triangulated and normals are generated. Finally, it is transferred back to the frontend application for rendering. Since multiple sub-volumes can be defined on the frontend, union \cup operations are also possible.

2.3 Level-of-detail rendering

The frontend application has two main tasks. On the one hand it has to maintain the ‘cut’; this is a set of nodes which are currently rendered on the screen (see Fig. 3). For all changes of the cut, a message is sent to the backend. On the other hand, it has to perform the actual rendering of nodes for which data was received.

At first, only the root node is part of the cut. If it is visible and its bounding box covers a maximum solid angle, a split operation is performed: All visible children nodes are added to the cut and the root node is removed. This process continues recursively until all visible nodes have been refined according to the threshold angle.

Every frame this traversal is repeated, either adding nodes to the cut (because they became visible or due to a split operation) or removing nodes which either became invisible or which have been merged in a collapse operation. Each node added to the cut gets a custom priority, determined by the distance between the virtual camera and the center of the node’s bounding box. Therefore, nodes close to the camera have a higher priority.

Then, three lists of nodes are assembled: One contains all newly visible nodes, one all nodes which are not visible anymore and one all nodes with changed priorities. These lists are sent to the backend which then starts the feature extraction for all new nodes, updates the priorities of scheduled requests and removes all obsolete requests.

When a parameter of the visualization is changed (e.g. the list of visualized sub-volumes or the current time step) all running requests are considered obsolete and for each node in the cut a completely new request is sent to the backend. The backend will then abort all pending requests.

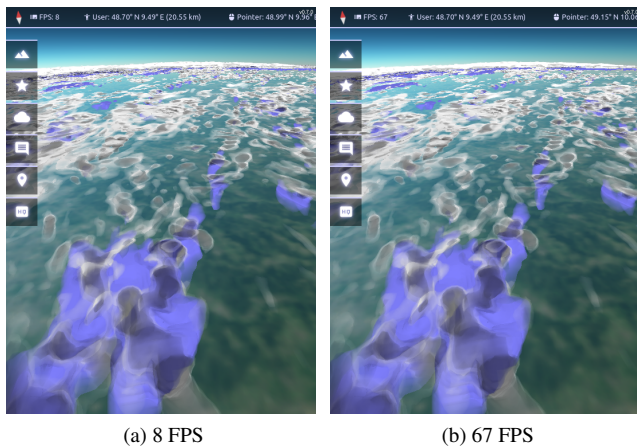


Figure 4: The full dataset cannot be rendered at interactive frame rates (a). However, this is possible if a view dependent resolution is used in the distance. The error in the distance is barely noticeable.

Rendering of the visible nodes follows a simple paradigm: A node which becomes visible should always show up-to-date data (the currently configured sub-volumes and the correct time step). However, a node which has been visible before is allowed to show stale data until it either receives new data from the backend or it becomes invisible. This results in a desired characteristic: When the configured sub-volume or the time step are changed, the scene progressively updates, starting with the nodes close to the camera. When frequent visualization changes occur, previous tasks are aborted. Hence, if the backend is not fast enough, there will be frequent updates close to the camera and less updates to distant nodes.

Optionally, the wind field is used to move the sub-volumes between time steps. This helps to make the transitions between time steps much smoother and conveys the motion of clouds.

In order to allow for the desired pseudo-volumetric appearance, the extracted geometry is rendered to an A-Buffer [3]. Usually, this data structure stores linked lists of semi-transparent fragments at each pixel which are filled during rasterization. In a final resolve step the lists are depth-sorted and alpha-composited. In STRIELAD however, no alpha value is written into the A-Buffer. Instead a unique sub-volume ID is stored for each fragment. This allows for resolving corresponding entry and exit points of arbitrarily intersecting sub-volumes along the view ray.

The color written into the A-Buffer is calculated by a user-defined function. This gets as input the sub-volume ID, all available interpolated scalar values at the corresponding point in the volume and some additional information such as the surface normal and the direction of the sun. A second function is invoked in the A-Buffer resolve step. It returns a transparency value based on the sub-volume ID and the distance between entry and exit points into the volume.

3 RESULTS

The developed toolkit can be used for real-time analysis of arbitrarily large structured atmospheric datasets in time and space.

With the level-of-detail rendering approach, all time steps of the entire dataset can be explored at interactive frame rates (see Fig. 4). When stepping through time, updating the current perspective takes only a couple of seconds. Since nodes in the foreground will receive data in less than a second this results almost in a fluent animation. This is especially true if the wind field is used to move the clouds between time steps. Furthermore, as the scene is updated progressively, the user can move the camera during the update with the full frame rate.

The user can specify multiple sub-volumes bounded by arbitrary scalar thresholds. This can be used to identify volumes of interest such as global extrema, trends and correlations between different scalars. Complex boolean operations can be performed on the dataset. The user can specify color and opacity for all sub-volumes with a powerful shader editor which allows for many different scenarios. The color can be based on all available scalars, their respective minimum and maximum values, the surface normal, the sun direction, the position in 3D space and other parameters.

Additionally, various digital elevation models (e.g. SRTM30) can be used together with other GIS sources such as OpenStreetMap or the BlueMarble dataset by NASA. The position of Earth and Sun is calculated with the SPICE library [1] and represents the conditions at the simulation time. Finally, the flight trajectories provided for the visualization contest are shown as lines with a user defined length.

4 DISCUSSION AND FUTURE WORK

With STRIELAD, we presented a scalable and interactive visualization environment for the analysis of large time-dependent weather simulations in combination with geospatial data. A level-of-detail rendering and processing approach combined with distributed and parallel feature extraction enables the real-time exploration of data. However, by now the feature extraction is limited to the selection of sub-volumes using boolean operators with powerful styling capabilities due to the shader editor. The integration of further algorithms such as slicing or contouring will be straight forward. More challenging will be algorithms using global information such as topological features or tracing methods. For these algorithms several aspects of the parallelization and data distribution will need substantial changes. Furthermore, our approach produces some artifacts when the maximum number of fragments in the A-Buffer is reached. For this purpose different fragment inserting strategies are necessary.

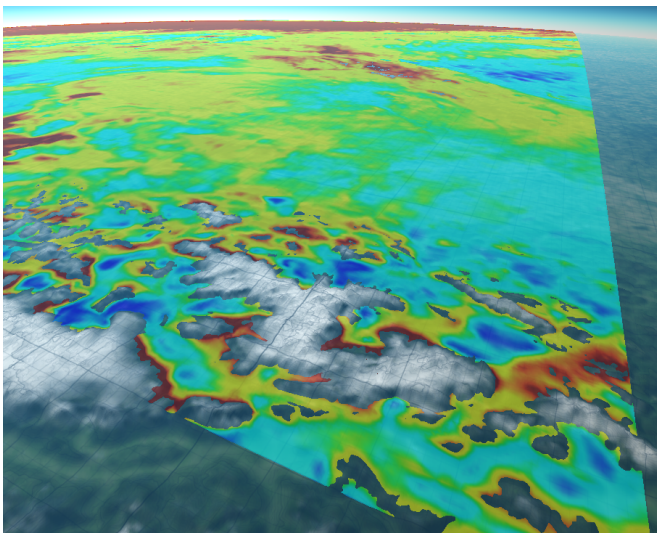
In the near future, we plan to add interpolation methods to the pre-processing. This should lead to much smoother transitions between the resolution levels of the LOD data structure. Additionally, we will extend the application to run on our tiled virtual reality system. To further speedup the feature extraction, we will investigate in data prefetching strategies.

REFERENCES

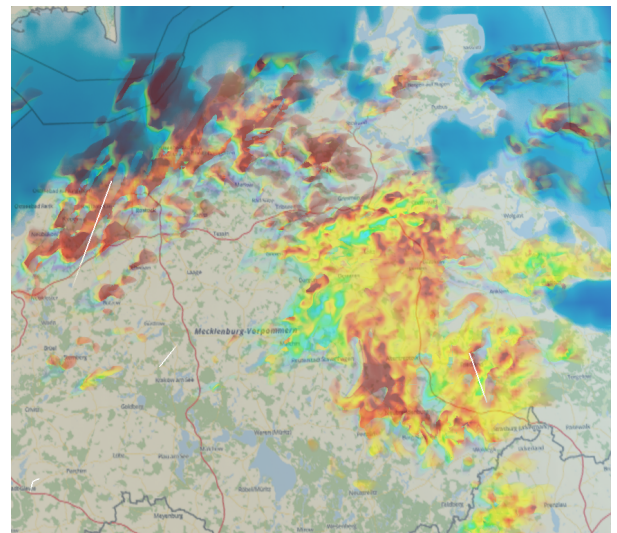
- [1] The spice toolkit. <https://naif.jpl.nasa.gov/naif/toolkit.html>. Accessed: 2017-07-27.
- [2] J. Ahrens, B. Geveci, and C. Law. Paraview: An end-user tool for large-data visualization. In C. D. Hansen and C. R. Johnson, eds., *Visualization Handbook*, pp. 717 – 731. Butterworth-Heinemann, Burlington, 2005. doi: 10.1016/B978-012387582-2/50038-1
- [3] L. Carpenter. The a-buffer, an antialiased hidden surface method. *ACM Siggraph Computer Graphics*, 18(3):103–108, 1984.
- [4] H. Childs, E. Brugger, B. Whitlock, J. Meredith, S. Ahern, D. Pugmire, K. Biagas, M. Miller, C. Harrison, G. H. Weber, H. Krishnan, T. Fogal, A. Sanderson, C. Garth, E. W. Bethel, D. Camp, O. Rübel, M. Durant, J. M. Favre, and P. Navrátil. VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. In *High Performance Visualization—Enabling Extreme-Scale Scientific Insight*, pp. 357–372. Oct 2012.
- [5] A. Gerndt, B. Hentschel, M. Wolter, T. Kühlen, and C. Bischof. Viracocha: An efficient parallelization framework for large-scale cfd post-processing in virtual environments. In *Supercomputing, 2004. Proceedings of the ACM/IEEE SC2004 Conference*, pp. 50–50, Nov 2004. doi: 10.1109/SC.2004.66
- [6] Kitware, Inc. *The Visualization Toolkit User’s Guide*, January 2003.
- [7] R. Westerteiger, A. Gerndt, and B. Hamann. Spherical terrain rendering using the hierarchical healpix grid. In *OASIS-OpenAccess Series in Informatics*, vol. 27. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012.



Figure 5: An airplane landing at Hamburg Airport. All 27000 airplanes are loaded and populate the scene with their trajectories. The length of the trajectories can be adjusted and defaults to the distance the respective airplane traveled in two minutes



(a) Visualization of color coded absolute wind speed in the south east of the domain. Interactions between the wind magnitude and the SRTM30 terrain are clearly visible.



(b) Sub-volumes of high rain mixing ratio color coded by the wind magnitude at their surfaces. Red areas indicate a lot of rain with high wind speeds.

Figure 6: The sub-volumes can be colored with complex shader code. In these examples scalars are mapped on a color scale.

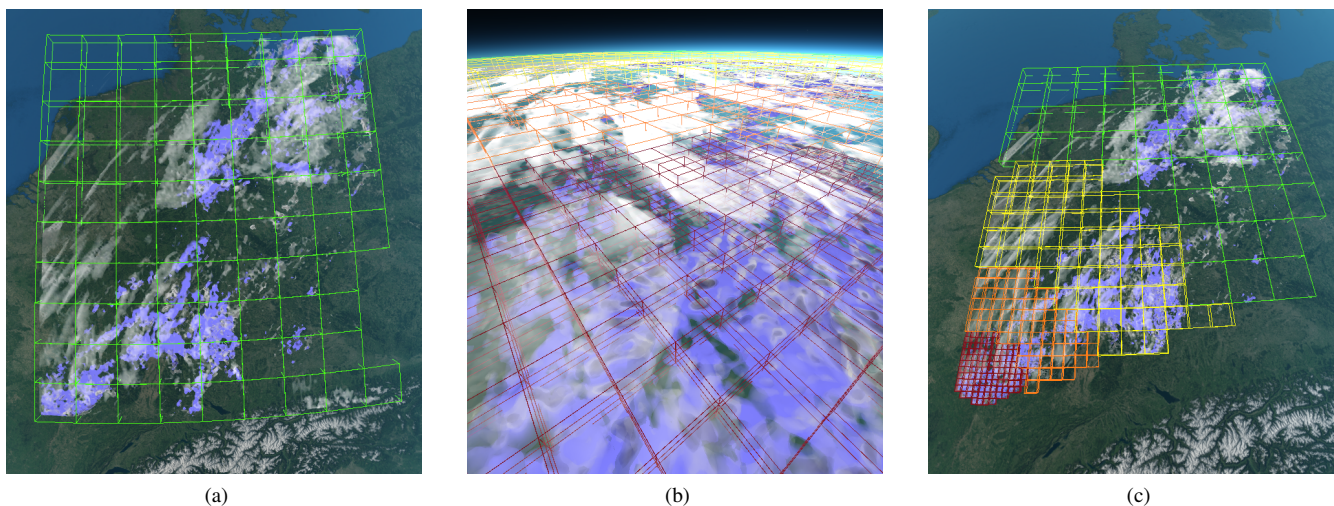


Figure 7: When the dataset is rendered for perspective (a), medium resolution nodes of the octree are chosen (illustrated with the green bounding boxes). If the camera is moved closer to the dataset, nodes with better resolution are loaded close to the camera (b). Image (c) shows the same nodes as (b) but rendered from an outside perspective.

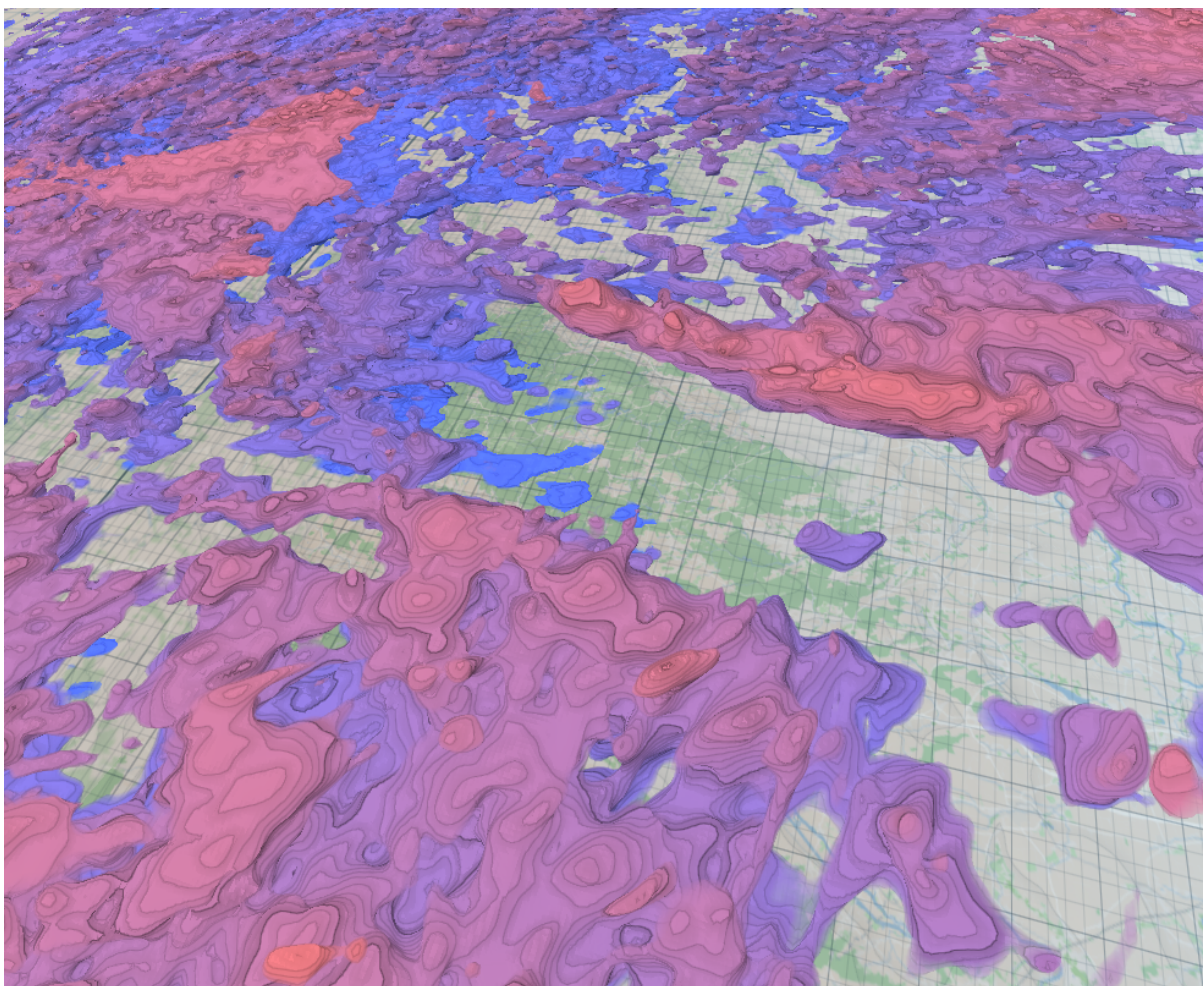


Figure 8: The shader editor allows expert users to achieve complex results. In this example, Sub-volumes of high cloud water content around the Harz are enriched with iso-altitude lines. Thin lines are spaced 100 m, thick lines 1 km in height. This helps to visualize the topology of the features.

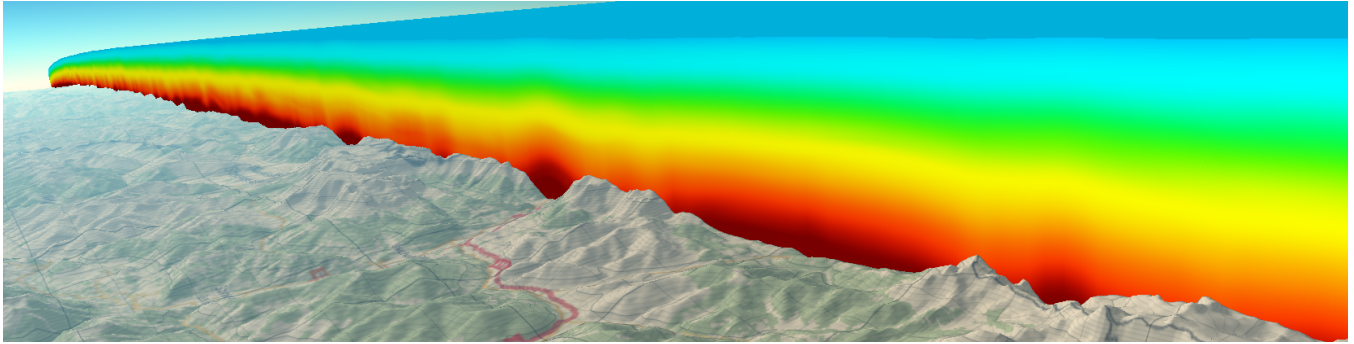


Figure 9: The southern border of the dataset with color-coded temperature shown with SRTM30 digital elevation model and OpenStreetMap data. The correlation between terrain and temperature distribution is apparent.



Figure 10: The visualization framework is integrated in an OpenGL application for rendering entire planetary datasets. In the screen shot above the relative position of Earth and Sun are as they were on 26/04/2013 at 5:03 PM. The tiny white dots are the airplane trajectories which were provided for the 2017 IEEE visualization contest.

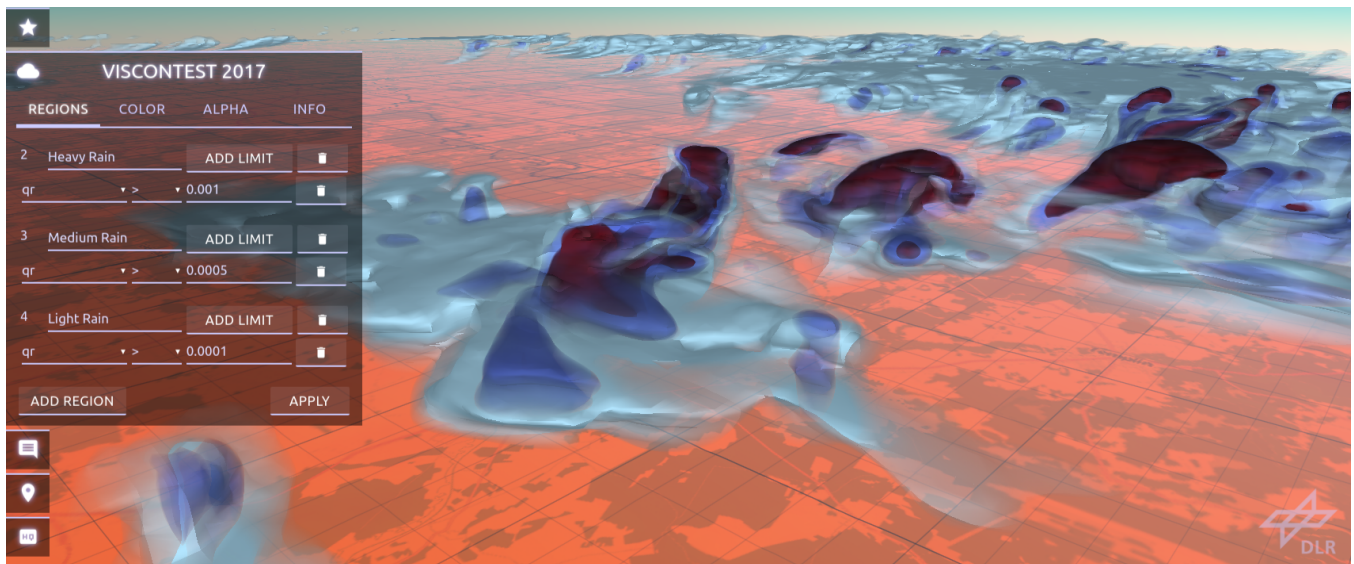


Figure 11: This is an example of intersecting volumes. Sub-volumes for three rain mixing ratios are inside each other. The pseudo-volumetric effect helps to understand the depth of the extracted regions.

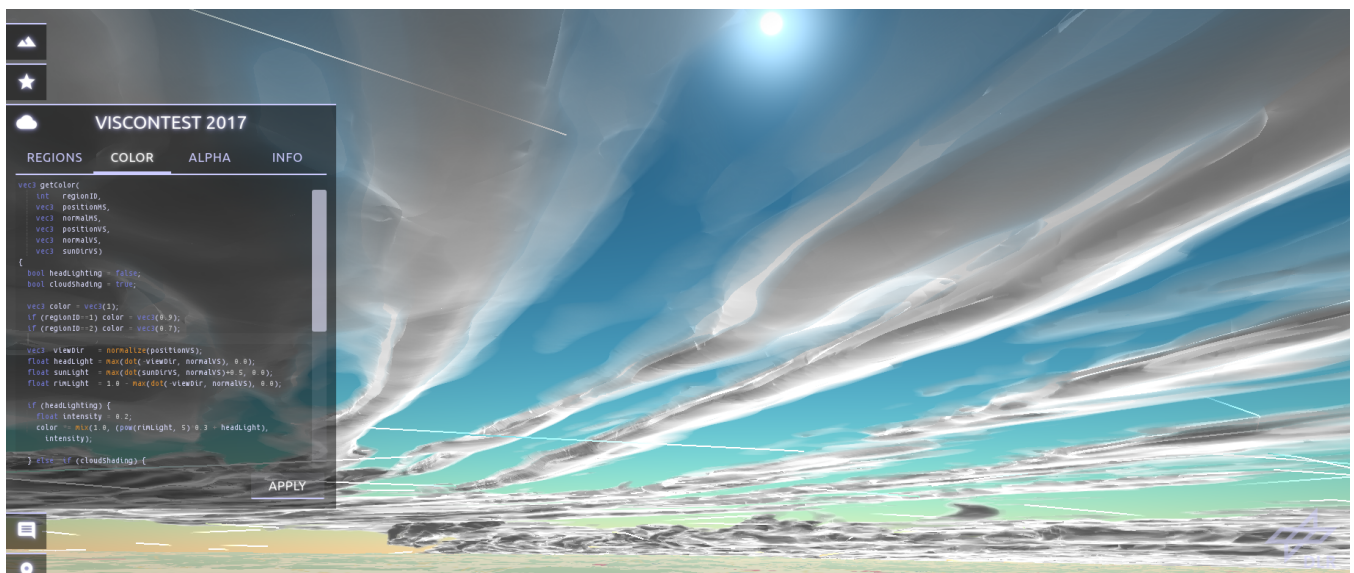


Figure 12: High altitude clouds in the southern part of the simulation domain. On the left hand side the shader editor in the user interface can be seen. There are also some airplane trajectories visible in this image.

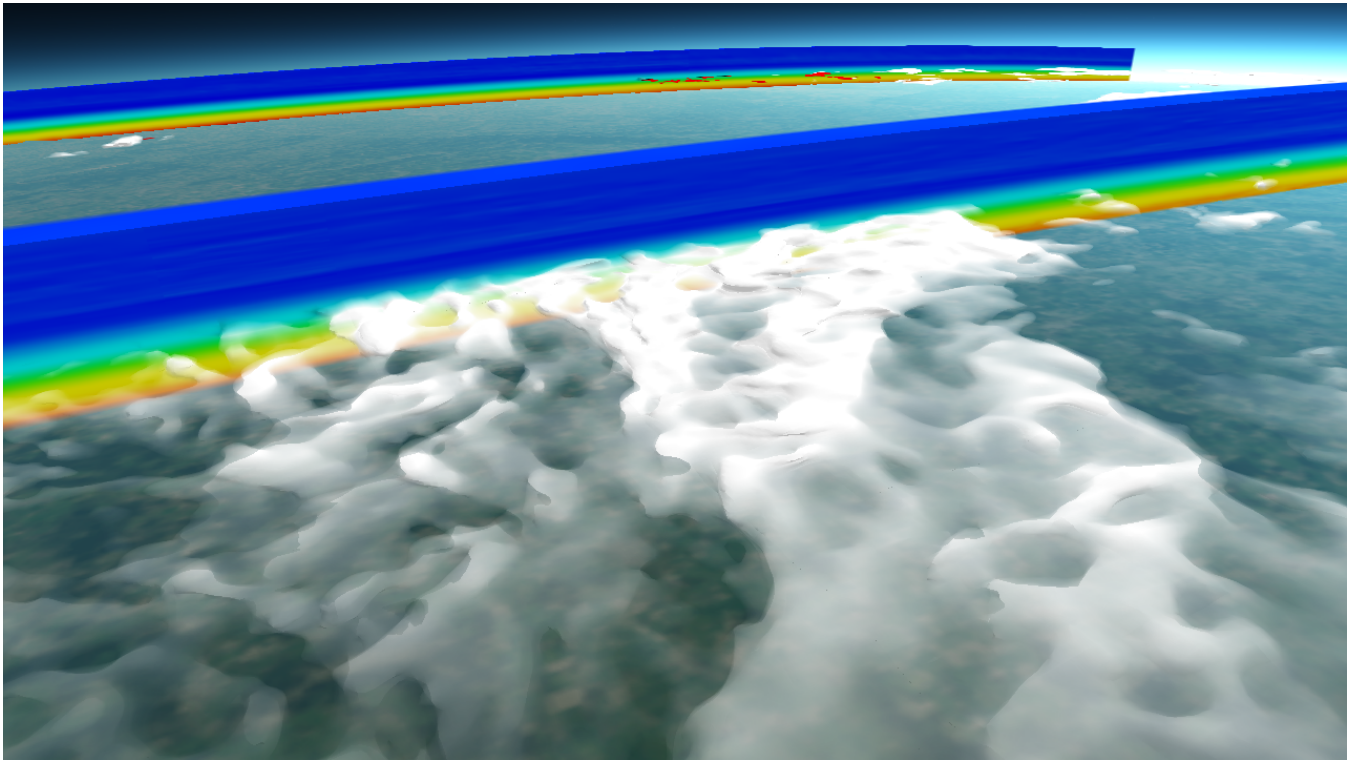


Figure 13: This visualization shows two slices of the dataset color-coded according to temperature. In addition, high altitude ice clouds are shown in white. None of the ice clouds extends into the tropopause (the darkest blue indicates the coldest layers of air) which would indicate a severe thunderstorm.



Figure 14: If a slice of the volume is color-coded according to pressure, gravity waves become visible.